DOCUMENT RESUME

ED 293 880 TM 011 474

AUTHOR Baker, Eva L.; And Others
TITLE Report of State Level Activities: Guidelines for the RFP Process and Selected Technical Issues in Large-Scale Assessment Programs. Report of State Level Activities. Project: Monitoring the Implementation of Testing and Evaluation Innovations.
INSTITUTION Arizona State Univ., Tempe.; California Univ., Los Angeles. Center for the Study of Evaluation.; Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; Colorado Univ., Boulder.; National Opinion Research Center, Chicago, Ill.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 87
GRANT OERI-G-086-0003
NOTE 81p.
PUB TYPE Guides - Non-Classroom Use (055) -- Collected Works - Serials (022)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Administrative Policy; Contracts; *Educational Assessment; Educational Finance; Educational Innovation; *Evaluation Methods; *State Programs; Test Construction; *Testing Programs
IDENTIFIERS *Requests for Proposals

ABSTRACT

Guidelines for evaluating responses to requests for proposals (RFPs) for state-level educational assessment programs are presented. The guidelines articulate key choice points, options, and considerations by state (and district) testing directors as they solicit services for large-scale assessment from commercial testing companies. Focus is on broadening the scope of issues and concerns that state and district testing officers consider in preparing RFPs for large-scale testing programs. Components that must be considered in dealing with the RFP process include: (1) planning approaches; (2) communications with bidders; (3) the RFP structure; (4) the review process; and (5) issues associated with equating portions of RFPs, item bias portions of RFPs, and content validity portions of RFPs. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

C·R·E·S·S·T

Center for Research
on Evaluation, Standards,
and Student Testing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Lisa Siegel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED293880

Center for Research on Evaluation,
Standards and Student Testing

Deliverable - November 1987

Project: Monitoring the Implementation of Testing and Evaluation Innovations

Report of State Level Activities: Guidelines for the RFP Process and Selected Technical Issues in Large-scale Assessment Programs

Center for Research on Evaluation,
Standards and Student Testing

Deliverable - November 1987

Project: Monitoring the Implementation of Testing and Evaluation Innovations

Report of State Level Activities: Guidelines for the RFP Process and Selected Technical Issues in Large-scale Assessment Programs

Eva L. Baker

Pamela E. Aschbacher

Study Directors

Grant Number: OERI-G086-0003

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

The project presented, or reported herein, was performed pursuant to a Grant from the Office of Educational Research and Improvement/Department of Education (OERI/ED). However, the opinions expressed herein do not necessarily reflect the position or policy of the OERI/ED and no official endorsement by OERI/ED should be inferred.

GUIDELINES

FOR THE RFP PROCESS AND SELECTED TECHNICAL ISSUES IN LARGE-SCALE ASSESSMENT PROGRAMS

Pam Aschbacher, Eva Baker, and Joan Herman
UCLA Center for the Study of Evaluation

Richard Jaeger
University of North Carolina at Greensboro

H.D. Hoover
Iowa Test of Basic Skills
University of Iowa

Ron Hambleton
University of Massachusetts

TABLE OF CONTENTS

PREFACE i

INTRODUCTION 1

PART I

TOWARD A MORE EFFECTIVE REP PROCESS: ISSUES AND OPTIONS 7

A. BASIC ISSUES 8

B. APPROACHES TO PLANNING 17

C. COMMUNICATING WITH BIDDERS 20

D. RFP STRUCTURE 24

E. THE REVIEW PROCESS 34

PART II

TOWARD BETTER TECHNICAL QUALITY IN LARGE-SCALE ASSESSMENT PROGRAMS: ISSUES TO CONSIDER 41

A. ISSUES IN THE EQUATING PORTIONS OF RFPs 42

B. ISSUES IN THE ITEM BIAS PORTIONS OF RFPs 47

C. ISSUES IN THE CONTENT VALIDITY PORTIONS OF THE RFPs 57

PREFACE

These Guidelines were developed as part of the Model RFP Project conducted at the Center for Research on Evaluation, Standards, and Student Testing (CRESST) located in the Graduate School of Education at UCLA. The Project was part of a larger program titled Monitoring Improvement in Testing and Evaluation Innovations (MITEI), which was funded by the U.S. Office of Educational Research and Improvement (OERI).

The Model RFP Project was developed in collaboration with state testing directors as an approach to improve the technical quality of state tests that assess educational performance of students and teachers. To this end Project personnel conducted interviews and surveys, reviewed recent requests for proposals (RFPs) for state testing, and held a two day meeting at UCLA in May, 1987, which was attended by three representative groups: state testing directors experienced in the RFP process, commercial test companies who bid on such RFPs, and researchers from the academic measurement community.

The project's original objective was to develop model language for a state assessment RFP. However, during the course of the project's activities, an urgency for improving the entire RFP process was revealed, particularly among testing directors and vendors, and the focus of the project expanded.

At the Model RFP Project's May meeting, participants decided to address problems in the generic RFP process as well as issues in

specifying standards of technical quality in RFPs. Participants discussed various choice points and options in the RFP process and the treatment in RFPs of such technical concerns as equating, item bias, and content validity.

After a general discussion of RFP procedural problems and three technical issues, participants divided into two groups: one that focused on improving the RFP process, and one that focused on technical concerns.

The first group devoted its time to articulating choice points and options in the RFP process. Members of the group voiced concerns about a gamut of issues, including those related to fairness, quality assurance, cost, and communication of expectations. The group tried to take into consideration the differential constraints of state regulation and the competitive nature of the RFP process. Part I of this document summarizes the Guidelines developed by this group.

The second group had time to deal with only two of the three technical concerns that were discussed at a general level during the opening session. Since there was more controversy surrounding item bias and equating, these topics were selected for further attention. Members of this group agreed that the standards of technical quality for tests should be explicitly included in both the requirements section of RFPs and in the criteria for judging the vendors' proposals. However, members discovered that there was considerable disagreement about which approaches had greatest merit within each of the two technical areas. There was consensus that some

states have required too little of vendors to assure the technical quality of the tests. Other states have sometimes required inappropriate practices, such as asking vendors for equating studies with expectations far beyond what measurement experts believe to be psychometrically sound practice. Because of legitimate and significant disagreements over preferred technical strategies and because of difficulties in anticipating specific data conditions, this group was unable to provide step-by-step directions to states or model RFP language. Instead, the group felt it appropriate to encourage the development of RFPs that require vendors to identify decision rules that should be used at critical choice points and to be as specific as possible in stating and justifying their technical approach.

In addition, members of the group felt it important for states to be aware of the experts' methodological disagreements before developing RFPs, evaluating proposals, and contracting for technical services. The papers on item bias and equating in Part II of this document summarize the discussions of these technical issues. (The group did not address the issues of reliability, setting of passing scores, or the details of the test development process; this exclusion was a result of the constraints of time, and in no way suggests that these issues are of lesser importance.)

The short paper on Content Validity in Part II summarizes the somewhat briefer discussion of that topic by the entire group during an earlier session of the meeting. Although it is less comprehensive, we have included it because it will undoubtedly be of interest to

some readers. Content validity was recognized by the participants as a particularly important topic and one that we hope to address more fully in future.

The participants of the May meeting suggested two avenues for continuation of the MITEI Project. One is to append this document from time to time with additional sections, such as a model RFP outline, optional model RFP language, and position papers on technical issues such as reliability, item analysis, standard stetting, construct validity, content validity, customized norms, and test score reporting. A second plan is to create an independent review or standards committee for RFPs that provides such services as the following: allowing states and vendors to anonymously air concerns and provide feedback on one another's efforts, maintaining a collection of RFPs and model RFP language for technical issues, and maintaining a directory of "qualified bidders."

In addition to the authors, the following people participated in the May meeting and reviewed drafts of this document. Their generously provided expertise, time, and overall support for the project have been invaluable.

Joan Baron
William Brown
Leigh Burstein
John Keene
Thomas Kerins
Stephen Koffler
Elaine Lindheim
Wayne Neuberger
W. James Popham
Edward Roeber
Paul Sandifer
Ramsay Selden
Stephanie Zimmermann

We particularly appreciate the efforts of Richard Jaeger, H.D. Hoover, and Ron Hambleton, who pulled together the thoughts on Equating, Item Bias, and Content Validity that were generated during meeting sessions and who authored those sections of the Guidelines. We also thank Lawrence Rudner and Tom Fisher for their helpful comments and suggestions.

Although these guidelines specifically address the RFP process, we recognize that many states and districts use other procedures for obtaining their tests, including in-house development and development through special relationships with state universities or other local organizations. We hope that our guidelines will prove of some use in these settings as well. Despite the fact that we originally intended the audience of our project to be state testing directors, we hope that our efforts will also be useful to district testing personnel.

P. A.

INTRODUCTION

Purpose

These guidelines articulate key choice points, options, and considerations faced by state (and district) testing directors as they solicit services for large-scale assessment from commercial testing companies. Since so many states and districts write Requests for Proposals (RFPs) to procure services from test service vendors, we have focused these guidelines on the RFP process.

The questions and discussion presented in the Guidelines are intended to broaden the scope of issues and concerns which state and district testing officers consider in preparing RFPs for large-scale testing programs. Many of the same issues and problems are faced by states and districts that use methods other than RFPs to plan and obtain testing services, so we expect the Guidelines to be of some use to them as well.

Prior to our work on this project, many of the issues described here had not been discussed openly among state testing directors and vendors. The frank discussions that developed among experienced testing directors, representatives of major commercial test companies, and measurement specialists during project sessions highlighted ideas and explicated options that should be useful to others involved in procuring large-scale assessment services. We hope that this effort will be particularly useful to new testing officers and those not part of a supportive network. Commercial test vendors may also be interested in these guidelines.

We have tried to present a logical, straightforward approach to planning and developing RFPs for large-scale assessment. Our approach has been guided by the assumption that one of the primary goals of an effective RFP process is to obtain a reasonable number of proposals that are "on target," that is, which speak directly to the needs of the state or district. However, there are significant tensions inherent in the nature of the RFP process that obstruct the establishment of a set of failsafe procedures. These tensions were referred to frequently during our project meeting to explain states' and vendors' behavior or motivation in a variety of circumstances. A brief overview here of the way these forces operate will delineate our own point of view in developing these materials and facilitate your implementation of these Guidelines.

Tensions in the RFP Process

The provision of testing services is big business, and this results in competing interests: constrained versus open communication, creative approaches versus use of carefully specified detail and adherence to RFP requirements, and cost versus quality. In addition, the limitations imposed by local or state policies and procedures inhibit implementation of an ideal, logical RFP process. Herein lies the root of many of the decisions made and the difficulties suffered by both testing directors and vendors.

In an attempt to preserve fairness in a very competitive atmosphere, states develop policies that sometimes result in very limited communication between state and bidders. For example, some states do not allow testing directors to talk directly with

bidders prior to the proposal deadline. All communication is routed through the purchasing agent. Many try to maintain fairness by states soliciting questions from bidders and then sending a written document of all questions and answers to every bidder before the deadline. While this practice would be expected to facilitate fair communication, it often fails to do so. Because of the intense competition, bidders tend not to ask significant questions that might reveal their approach to a problem or clarify for other bidders as well as themselves important aspects of the state's needs. Some bidders also hesitate to ask questions for fear of revealing their poor comprehension of the project at issue. In addition, by the time bidders receive the written answers, it may be too late to use this information to modify their proposals.

The communication problem is exacerbated by the fact that many RFPs are quite vaguely worded for any of several reasons. Sometimes RFP authors are uncertain about the purposes or details of a new program, especially during the early conceptual stages of the program. In some cases, testing directors are faced with such long timelines in getting RFPs authorized, reviewed and accepted by various state officers (up to a year in some cases) that they must dovetail tasks very closely in order to meet a deadline. This may result in their having to write the RFP before some of the essential ground work has been completed. For example, they may not be able to describe the number of objectives and items to be developed because the objectives committee has not finished composing the list.

Another cause of vaguely-worded RFPs is the desire to enhance competition as a means to obtain better test services for less money.

Some states have provided limited specificity to bidders, especially in terms of the level of effort and scope of work expected. The notion is that less information will spur the creativity and competitiveness of the bidders. The consensus of our meeting, however, was that this notion is a myth. Explicit information in an RFP about the expected cost and scope of work facilitates rather than inhibits good proposals.

Vagueness, on occasion, does yield a highly creative, reasonable bid. However, bidders' solutions to vaguely-stated requirements may be so diverse in scope, quality and cost that bids cannot be fairly compared. Furthermore, this situation may open the door to legal challenges of the bid process by one or more of the bidders.

Since cost is so important to states, budgetary concern often shapes the focus of the planning effort and the RFP itself, sometimes to the detriment of technical quality issues that need to be addressed. For example, much attention may be devoted to specifying the number of meetings to be held, where they will be held, who will attend, and how much is budgeted for lunch, but no specifications may be provided for the purpose of the meetings (e.g., how to establish the content validity of the test). If the RFP emphasizes process over purpose, the resulting tests may suffer in quality as a result.

Concern for costs may also directly affect district or state policy, impacting bidders and in turn the alternatives available to the state or district. If it is known that a state contract is likely to go to the lowest bidder, because of either official or practical policy, vendors with possibly better plans but higher costs often will not

bother to submit bids. States represented at our Model RFP Project meeting said their RFPs typically solicited bids from only three to four vendors. Given the size and importance of many projects, most states would prefer a greater choice of proposed services. If none of the proposals is adequate, additional time and money may have to be spent to issue a revised RFP, or the state may have to shelve the project.

Although states tend to think of competition as resulting in lower costs, they must also be aware that quality may be compromised. There are many reasons why a vendor, large or small, may provide a low price, and this may or may not benefit the state. For example, a test company may underbudget a proposal for an early test development effort in a state (taking a loss on that contract) to position itself favorable for the larger related testing contracts that may occur there in future. The first contract may yield a high-quality, low-cost test—a real bargain. However, the bargain may be balanced later by the cost of future contracts, either with the same vendor who must recover his costs eventually, or with another vendor whose costs may be higher because he was not involved in the original work on that program. In any case, the state should attempt to protect the quality of its programs by specifying its expectations for quality in the RFP, providing for reference checks for vendor performance, and carefully examining the consequences of its low-bid policies.

Organization

The Guidelines are presented in two major parts: one on planning and writing your RFP, and the second on technical issues that you may need to deal with in the RFP.

Part I deals with issues and options to consider for a more effective RFP process. The material is organized into five sections and is arranged in an order that follows the chronology of the RFP process.

The first section covers basic issues that should be considered prior to any RFP writing: the amount and type of money available, the amount and flexibility of time available, the degree to which the project calls for innovative approaches, whether the bid is to be competitive, and whether there will be one or more phases of RFPS required to accomplish the entire project.

The second section of Part I discusses the pros and cons of two methods that can be used to facilitate the planning of new programs before you write the RFP: concept papers and planning meetings. These methods involve obtaining expert advice on how to handle a new project or approach a particularly thorny technical issue.

The third section describes methods to facilitate bidders' understanding of what is required in the RFP: letters announcing upcoming RFPs, bidders' conferences, and methods of handling bidders' inquiries.

The fourth section discusses the major sections of an RFP and emphasizes the importance of articulating your needs and priorities to assure a top-quality and cost-effective product. Topics include the introduction of the RFP, expected cost of contract, the scope of work, technical design and report, expected services and products,

personnel loadings, the RFP budget, and quality control and scheduling.

The fifth and last section of Part I is about structuring the review process and pre-contract negotiations to assure an economical product of the highest quality. Included here are discussions of specification of criteria and process, relative weights of technical quality and cost, members of the review panel and their qualifications, usefulness of oral presentations, and factors to negotiate before the contract is finally let.

Part II of this document consists of three papers on issues and specifications in equating, item bias, and content validity. The latter paper summarizes a brief presentation on content validity at the MITEI Project's May meeting and the participants' general discussion of the topic. The equating and item bias papers present a summary of the more focused discussion that occurred on these two technical issues during the second half of the May meeting. Each paper presents an overview and suggested specifications for RFPs dealing with these technical concerns. Future additions to this part of the document are planned.

PART I

TOWARD A MORE EFFECTIVE RFP PROCESS: ISSUES AND OPTIONS

A. Basic Issues

B. Approaches to Planning

C. Communicating with Bidders

D. RFP Structure

E. Review Process

A. BASIC ISSUES

This section considers five fundamental aspects of the testing project that will significantly influence the planning of the RFP process and document:

1. Money
2. Time
3. Conventional or Innovative Project
4. Type of Bid
5. Single vs. Multiple Phases

1. MONEY

How much money is available for the testing project? Is the money fixed, flexible, or a combination?

The amount of money available for a project is a critical factor in what the project can hope to accomplish and thus in the scope of work set forth in the RFP, so test directors have a responsibility to budget well before issuing an RFP. It is important to consider whether funding has been granted for the entire project or only for a portion of it. The amount of funding and how it is scheduled will influence the number of RFPs that you will write for a project. Such information will also be important to communicate to bidders and should be included in the introductory section of the RFP itself, if that is allowed in your state (see section D-2). In addition, certain contracts can be awarded in whole or in part.

Whether the money is fixed, variable, or a combination will also influence how you write the RFP. If the money is variable and the final cost of the project turns out to be more than expected, you may have to deal with the added effort, expense, and bother of rewriting the contract. One strategy to avoid rewriting when actual costs are unknown is to estimate the biggest possible number of students to be tested, materials to be printed, and so forth. Then the needed money will be available without having to rewrite the contract. Contracts should have a provision that payment will be for work done. For example, if the RFP calls for budgeting on the basis of 100,000 students, and only 85,000 are tested, the vendor's bill for that work ought to reflect the 85,000 students.

For many projects you may want to include both some fixed costs and some variable ones, such as scoring costs per student or optional reports paid for by the state that individual schools might elect to receive. In the end, the cost to the state will be the fixed costs plus the variable costs minus any credits (given by the vendor for contracted work that you mutually agreed to omit in return for a credit) and minus any penalties (such as those assessed for late delivery of materials or services). Requests for Proposals can have an options section with items to be included should funds become available. This can eliminate the need for new procurements.

2. TIME

How much time is available? Is there any way to increase the flexibility of time schedules?

There are several major constraints that may affect the time available for a project and its degree of flexibility: legislative mandates, funding tied to the fiscal year, contingency of part of a new project on previous work having been completed, and the time required by purchasing agents and others with whom you must work to let the contract. These constraints will affect your plans for completion of the project and will determine a portion of the information communicated to the bidders prior to issuing the RFP and in the RFP itself.

When a project is mandated by the state legislature, the timeline is usually non-negotiable. Your best bet, where feasible, is to try to influence the generation of the legislation before it is actually passed. Since mandates differ in their level of prescription, it is obviously to your advantage to encourage a less prescriptive mandate in which you may be able to set up at least some of your own timelines and may be able to deviate from them when it proves necessary. Failing that, you can attempt to work with vendors to educate legislators, governors, and their aides regarding what sort of timeline would be minimally adequate for accomplishing a high quality project. (CRESST hopes to address this problem in the near future by including policymakers in the dialogue for improving the nature of large-scale assessments and by providing states with

materials or other means to communicate to policymakers the importance of quality in large-scale assessments, and the importance of having adequate time and resources to achieve the desired quality.)

The contract dates for some testing programs are set by the state's Department of Education (DOE), and are usually tied to the fiscal year. In this case you must help the DOE set reasonable timelines with sufficient flexibility before you write the RFP. Scheduling flexibility is further enhanced by the ability to carry over funds from one year to the next. Project schedule, cost and quality are in a delicate balance, and it is important that contracting agencies and vendor organizations be aware of this. When schedules are compressed, costs may increase because of the use of additional staff, overtime pay, courier services, etc., and the number of quality assurance steps may be reduced or eliminated.

Large scale assessments rarely exist in isolation. Such testing programs sometimes resemble a very complex puzzle comprised of many small parts that must be integrated in terms of time. When writing an RFP for part of such a complex situation, it is important that you consider what work may need to be accomplished before following portions can be done and then structure the timelines and RFPs accordingly.

In some cases, a great deal of time must be budgeted to shepherd the RFP through a variety of required administrative

procedures, such as gaining authorization for the RFP (which may take up to eight months), writing and reviewing boilerplate sections of the RFP, scheduling when the RFP will be issued, scheduling a pre-bid meeting, listing potential vendors, scheduling the review committee, preparing insurance forms, and so forth. Sometimes Commissioners of Education or Assistant Commissioners can help to expedite this process.

Time must also be budgeted for the bidders to respond to the RFP, and thoughtful responses require time. Typically, about four to six weeks are allowed, but two to four months (depending on project complexity) would be desirable. Scheduling pressures are often increased by deadlines, such as the need to start the project before the end of the fiscal year or the need to field the assessment by a certain time in the school year. Unfortunately, some of these tasks and deadlines may be outside the control of the testing director. In fact, you may need to work with purchasing agents, accountants, and others who know little about testing services.

In this situation it is imperative to anticipate these constraints and to organize and dovetail tasks to minimize negative effects on the program. In some cases you may be forced to use language in an RFP that is more general or vague than you would like simply because there is not sufficient time to wait until you know all the details. When there is little or no flexibility in your time schedule, you may be able to gain some flexibility by carefully wording the

RFP and fully using pre-RFP opportunities to communicate with vendors.

3. TYPE OF PROJECT

Is the testing project conventional or innovative? Does the technology for solving the problem exist or does it need to be invented? How committed is the state to a particular approach or specific solution?

There are three very different types of testing projects: (a) those in which the new project is to be an extension of an ongoing program or to replicate a model program; (b) those in which the new program is to differ significantly from what has been done in the past; and (c) those in which a new program is to be implemented where no program existed in the past. It is critical to tell the bidders which sort of program you want.

In the first case, in which a previous program or approach to a problem is to be replicated, it may be important to tightly specify what has been done in the past so that it can be repeated, right down to the number of items and the number of test booklets. If you are modeling a program after one used in another state, it is particularly useful to describe the similarities and differences between the two programs. Testing directors need to be careful, however, not to imply that a project is more complicated than it really is. The more

complex a project appears, the more vendors tend to budget for it. A misleading description may result in the state being overcharged.

In the second and third cases, where significant change or innovation is called for, you should be as specific as possible about where the innovation is desired. If you want a creative approach in one area, such as type of test item, but are committed to a particular approach or solution in another area, such as type of analysis, it is imperative to communicate that to the vendors. It is also important to carefully state the criteria for judging proposals, distinguishing between what is "required" and what is "desired but not necessary." Remember that if the review process goes by the numbers, requiring a creative approach means eliminating proposals that may have good but not "creative" approaches.

It is also essential to outline for the bidders any givens, decisions, or constraints within which creative solutions must work. This should include a detailed description of any previous related projects, particularly those aspects which should be avoided and those which might offer clues to success in new approaches. Validation and development costs will probably be higher with innovative projects, and implementation will probably take longer.

Many states include language in their RFPs stating that all ideas in the submitted proposals become the property of the state. States may then use any good ideas in the proposals without having to contract with the vendors who proposed them. While this outcome

may be useful to the state, the provision may restrict the expression of good ideas in proposals and discourage some vendors from bidding at all. Since certain states require that all proposals become the property of the state, placing them in the public domain, you may have no control over this situation.

4. SOLE SOURCE OR COMPETITIVE BID

Should the bid be sole source or competitive? Are there a number of vendors who could do the project well or really only one?

For most large-scale assessment programs there are many vendors who might be able to do the project well, and states are usually well served by the competitive bid process. In fact, it often would be premature for the state testing director to decide that only one vendor could or should have the contract. In some states, directors do not have this option. However, it is occasionally quite obvious that only one vendor (who possibly is subcontracting part of the work) is in a position to do the desired project, and then it is better and more efficient to work directly with that company if possible. Other vendors will not lose the time and resources involved in making a bid that would not have been seriously considered.

In some states you may not have to even write an RFP if there is a sole source for the project. In the case where a sole source must still submit a bid, you may be able to help them put their bid together, which will improve the eventual contract.

Some states feel that continuity with a single contractor over several years of a continuing program is important. In fact, the RFP process requires so much internal effort that at least one state is moving toward more five-year RFPs. However, one testing director recommends one- or two-year contracts in the beginning to avoid trouble through lack of experience.

5. SINGLE VS. MULTIPLE PHASES

Should there be one or more phases of RFPs to accomplish the project?

In some cases a single RFP will be all that is necessary to solicit the work to accomplish a program. However, many programs may be better served by multiple phases of RFPs. For example, a very large, complex, or innovative program may be best served by a multiple phase proposal process in which later development or implementation is dependent on an effective design or prototype produced during the first phase of the project. Each phase of the project may differ considerably in the amount of time and money available, the degree or type of innovation desired, and its suitability for sole source or competitive bids.

Sometimes you may not be able to find the type or quality of work you want for all aspects of a project at an affordable price from a single vendor. In this case, you may want to break the project into

parts, each part to be done by a different company. For example, one vendor might develop the test and provide a camera-ready copy and a second vendor might provide shipping, printing, data analysis and reporting. Dividing a program between two or more vendors makes the specification of responsibilities and timelines of each part of the program critical. It is also critical to state in the RFP which tasks may be split among vendors, since bidders often base costs, quality, and schedule on integrated processes.

Dividing the RFP into multiple phases can be problematic. Coordination strategies must be put in place to integrate multiple contingencies, monitor compatibilities, and prevent critical aspects of the project from being overlooked. The more people involved in the project, the more deadtime is necessary at the beginning to establish mandatory coordination. Geographic separation of the companies usually makes coordination more difficult, expensive, and time consuming. In addition each separate contract multiplies the red tape and amount of time required to develop the RFP itself. To minimize some of these problems, you can encourage the main vendor to subcontract.

B. APPROACHES TO PLANNING

This portion of the Guidelines discusses a couple of methods to assist the planning of new programs prior to writing the RFP, including:

1. Concept Papers or Individual Comments
2. Planning Meeting

1. CONCEPT PAPERS OR INDIVIDUAL COMMENTS

Should there be some sort of pre-RFP communication among state testing officers and others (such as measurement specialists and vendors) that invites concept papers about proposed testing plans or individual comments on a rough draft of the RFP? Should an outside consultant be hired to help with the RFP? How can inequities in distribution or in competitive advantage be avoided?

When considering a significant new program or new technical approach, you may want to gather expert input while conceptualizing the project, before finalizing the RFP. Unfortunately, this choice is often precluded by lack of time. When time permits, input from measurement specialists and vendors may be quite useful. You could consult them about state-of-the-art technical approaches to such areas as bias, validity, equating, or the scoring of writing samples. However, some vendors may be reluctant to give away their good ideas.

Pre-RFP concept papers or requests for individual comments could also be used as an initiation to the bidding process to identify qualified bidders. You could send your ideas for a new type of test to a number of vendors for their suggestions and comments, then invite some of them to respond to your RFP if you liked their initial response. To avoid charges of unfairness, the criteria used to select the initial grouping of responses should be specified ahead of time. Note that in some states it may be illegal to request pre-bid concept papers or invite only some vendors to respond to an RFP.

When requesting papers or comments, you should be open about (a) who can respond with the first comments or concept papers, (b) whether the ideas expressed in the comments or concept papers will belong to the state to possibly use in its ensuing RFPs, and (c) whether the eventual RFP bidding will be open to anyone or only to those who have been selected as "qualified" during the comments phase.

2. PLANNING MEETING

Should the state hold a planning meeting with vendors or test experts on how best to handle a new project or a thorny technical issue? Will the state pay for the cost of the meeting?

Another approach to gathering expert input prior to writing the RFP is a planning meeting involving measurement specialists and/or vendors. Group interaction can evoke many perspectives and

provide insightful solutions; however, don't expect consensus. A meeting will usually raise more questions than it answers, but raising these questions early may preclude major problems later in the process.

Measurement experts may be able to provide useful ideas on solving problems, conceptualizing the issues, and recommending procedures, but they need to be people familiar with the complexities of real world testing, not just academic ideas. Vendors also may be able to provide good insights about the problem or issue, but they may not wish to share these ideas at meetings where competitors are present. A written response, such as a concept paper, may suit them better. However, the testing project under consideration would have to be quite major to induce vendors to spend this much time on it without any assurance that they would get the final contract.

C. COMMUNICATING WITH BIDDERS

This section deals with methods of informing bidders about the RFP to enable them to fully understand what is required. The methods are:

1. Letter to Announce Upcoming RFP
2. Bidders' Conference
3. Bidders' Inquiries

1. LETTER TO ANNOUNCE UPCOMING RFP

Should an introductory letter be used to announce an impending RFP?

An introductory letter sent about a month in advance of an RFP has a couple of advantages. Perhaps most important, it lengthens the response time to an RFP, effectively doubling it in many cases. This permits vendors more time to consider whether and how to respond to an RFP. It allows them time to plan ahead for possible staffing allocations, to shift priorities, and to organize their time, all of which is important, particularly for small vendors. An introductory letter also provides a crucial period when vendors can ask clarifying questions in those states where communication is virtually cut off (other than through a purchasing officer) once the RFP is issued.

1. BIDDERS' CONFERENCE

Should a bidders' conference be held? Should attendance be mandatory?

A conference theoretically informs vendors about the state's needs, priorities, and commitments. This knowledge would be particularly helpful to vendors when the RFP is not very specific on certain points or when the project is expensive, complex or very innovative. The resulting proposals should more likely be on target.

Few conferences, however, are actually as useful as they could be. A bidder's main reason for attending is often to see who else is bidding. The usefulness of a conference in clarifying important details of an RFP tends to be limited by the competitiveness of vendors. They tend to be very guarded about the types of questions they ask to avoid giving away information to their competitors. Some meetings have been as short as ten minutes because no one wanted to ask any questions. Nonetheless, conferences can serve a useful purpose. Regardless of how routine the testing project might be, new vendors may want to bid, past years' practices may be changed, and so forth, and the meeting can be the source of important information.

Attendance at conferences may be mandatory or optional. A mandatory conference may help a state to weed out vendors who are not interested enough to send someone to a required meeting. However, the cost to vendors of sending staff to a conference,

particularly one far away, must be recovered by future business. Thus, the states, as consumers, will inevitably pay for any vendors' costs associated with such conferences. Some small vendors may simply opt not to respond to RFPs that require attendance at such meetings, so the effect may be to limit the number of vendors who send proposals to a state that uses this procedure. Some states hold conferences at which attendance is optional, allowing the vendor to weigh the advantages and disadvantages of participating.

Taping a conference is recommended for several reasons. The taped record can clear up misunderstandings, help reviewers, and protect the state from court action if necessary.

If the state knows exactly what it wants (such as continuing an ongoing project), and what it wants is fairly routine, a meeting is probably a waste of time and money for both vendor and state as long as the RFP is quite explicit.

3. BIDDER INQUIRIES

How can bidders' inquiries be fairly handled?

States handle bidder inquiries in a variety of ways, depending in part on state regulations. Their strategies range along a continuum from forbidding the testing officer to talk to any bidders during the RFP period to allowing any and all communication between a bidder and the testing officer at any time. One middle ground approach is to

send all bidders a written list of all questions raised and the answers before the proposal deadline. Each approach has its advocates and its advantages and disadvantages. Three illustrative approaches are described below:

Approach 1: All questions are referred to the purchasing officer; the testing officer is not allowed to talk directly to any bidders during the RFP period.

The purchasing officer should be a conduit, receiving calls from vendors, passing them on to the test director and relaying the answers. In some circumstances the purchasing officer may try to shorten the circuit and answer testing questions himself. He may not be knowledgeable enough to answer testing questions, and if he does not seek out answers in a timely fashion, the bidders will be left in the dark. However, this approach leaves all bidders in the same boat, so it is "fair" to all.

Approach 2: Questions may be asked, sometimes only in writing, and all answers are sent in writing to all bidders before the proposal deadline.

This approach is fair to all bidders; however, vendors may ask few substantial questions in order to avoid cueing competitors about their plans or the state's needs. In this approach, it is the test director's responsibility to minimize turnaround time, allowing

bidders the opportunity to use the information sent to them before proposals are due.

Approach 3: The testing officer can talk to anyone at any time (and may or may not send out written answers to all).

This approach leaves room for favoritism, particularly if written answers are not sent to all bidders. In this case the testing director should be careful not to give information to one vendor that would help him write a better proposal than anyone else. Proponents of this approach feel that it allows maximum communication between state and vendors, which may benefit both. It also rewards vendors who are savvy enough to ask good questions.

Opponents of this approach suggest that it is all too likely to result in litigation by vendors who don't get as much information as others. Even if all vendors do get the same information, the appearance of bias can have a very negative effect.

The importance of communication between state testing officers and vendors may be underscored by an anecdote. A vendor, hoping to win a contract by supplying a superior quality bid in a highly competitive situation, wanted to include some techniques for reducing item bias in his proposal although the state's RFP did not mention the subject. The vendor feared that if he included techniques for dealing with item bias in his proposal and the state did not carry them out for some reason, the state's failure to do so

might be held against it in court should there be litigation in the future (which was probable in this circumstance). The vendor's dilemma was whether to propose their best work and expose the state to some risk, or to do a lesser job, risk losing the contract, and not jeopardize the state. In this situation, communication between the vendor and state was critical to both parties.

D. RFP STRUCTURE

This section provides a description of the major portions of an RFP:

1. Introduction
2. Level of Effort
3. Scope of Work
4. Technical Design and Report
5. Expected Services and Products
6. Personnel Loadings
7. Budget
8. Quality Control and Scheduling

This section stresses the importance of structuring the RFP to articulate needs and priorities to assure a quality, cost-effective product. The proposal review portion of an RFP is discussed in Section E.

1. INTRODUCTION

Does the introduction clearly define the purpose or problem to be addressed by the RFP? Does it provide suitable detail on the programmatic context, background, and relevant legislation?

The introduction should contain several important pieces of information and provide the "flavor" of what you are trying to accomplish. First, provide a clear, concise statement of the testing program involved and the services and materials solicited. Be sure to specify whether you want to replicate a previous program or create

something totally different. Clarity cannot be overemphasized. For example, if you use the term "edit," do you want someone to redesign a test for you or just to make minor wording and format changes? This summary statement of purpose at the beginning makes it easier for vendors, who must read many RFPs, to quickly decide whether to respond.

Second, be as complete and accurate as possible in describing exactly what you want and what constraints, decisions, and commitments a contractor will have to deal with. When some important decisions have not been made yet, be explicit about what the decisions entail, when they are likely to be made, and who will make them (e.g., the state legislature, state board of education, state purchasing agent, state testing director). You may refer to attached documents that provide important information about what is to be accomplished, such as a paper on a particular design approach that is desired or a description of the new core curriculum for the state.

Third, discuss the "big picture," the context into which the proposed project will fit. Describe the relationship of the new project to other current or planned state programs, but do not bother the reader with description of unrelated programs. Information about related existing programs lets the bidders know that they do not have to address those problems in this project.

2. EXPECTED COST OF CONTRACT

Should or can the expected cost of the contract be made explicit?

Some RFP writers feel that not mentioning the expected cost of a project will result in lower cost bids and save the state some money. However, if vendors lack information about expected cost, they may propose approaches that are much too grand for the state's budget or just the reverse. If all proposals received are too expensive, the state is left in a difficult position. On the other hand, if a vendor underbids and wins, the state may be pressured into awarding the contract to someone who may not be prepared to do all that needs to be accomplished. As a result, the contract may have to be revised to accommodate additional expenses and the process may entail several months of lost time for the project. In addition, providing expected cost information helps vendors concentrate their efforts and resources on proposals for projects that they can ably handle.

In general, when expected costs are provided in the RFP, the state is more likely to obtain proposals that match its cost restrictions. Although the proposals may not match the state's desired technical quality, the state can focus its evaluation on the quality of the proposals in relation to what the expected budget can buy. The state can concentrate on getting the best product it can afford.

Specifying the probable cost of a project may reduce the number of proposals that state receives by weeding out the more expensive ones. Unfortunately, this may deprive you of seeing some good ideas and receiving feedback on your costing and scheduling, which may lead you to underestimate the effort actually needed to accomplish projects. Accurate projections of costs are especially difficult for new testing officers who must work with short timelines, low budgets, and high expectations. Feedback on costs and schedules can provide insight that helps these officers revise impractical estimates.

If your state law prohibits the provision of even ballpark figures, you may be able to provide relevant portions of a similar contract from a previous year, including the cost, which will probably be public information. This is particularly useful with large projects. Another possible strategy is to specify the expected cost in terms of the expected number of hours the project will need for completion.

If you are unsure how much of a flexible budget may be eventually allocated to a particular project, it is helpful to the vendor for you to be straightforward about the situation. Perhaps you can give high- and low-end figures. Vendors always have the option of bidding beneath the low end if they wish.

3. SCOPE OF WORK

Does the level of detail specified in the required scope of work match your understanding of task requirements?

When you know exactly what you want, say so. As simple as it sounds, this dictum is not universally followed. It is unlikely that vendors will be able to intuit expectations precisely. Furthermore, if a task is not mentioned in the RFP and thus is not a part of the contract, the state cannot compel the vendor to do it. Products, particularly, need to be specified as carefully as possible. Technical processes may be specified in detail if you are sure they are technically sound, but it is best to clearly state that bidders may suggest improved methods. This would allow vendors to use newer and better technical approaches with which you may not be familiar.

When you do not know what you want (which may happen in the early stages of a developmental project) it may be hard to be very precise. Loosely worded RFPs are usually intended to evoke vendors' creativity, but such RFPs often result in a group of proposals that differ so significantly that they are difficult to compare.

Vagueness, especially when occurring in the RFPs of a state without free communication between testing officers and vendor, can make the proposal writing very difficult for the vendors since they have little way of discovering what the state wants. Some good vendors may decide not to respond to such unclear RFPs, and the

state may lose the advantage of the vendor's competition and good ideas.

When you are in the early developmental or conceptual stages of a project, you are better served by:

(a) a pre-RFP planning meeting with consultants and possibly with potential vendors, to help clarify what you want so that the RFP can be precise and detailed, or

(b) a planning or design RFP, which can help you decide how to proceed with writing an RFP for the actual test development and other required services.

4. TECHNICAL DESIGN AND REPORT

Does the RFP require that vendors completely specify and justify all major elements of their technical design? Does the RFP include specifications for a full technical report? If specific technical requirements are included in the RFP, are they technically sound?

There are several reasons why many state testing officers prefer to allow bidders to propose their own technical methods rather than requiring specific techniques. At the time the RFP is written, you may not know exactly what you want in terms of technical design and analyses or if what you want is technically sound and state-of-the-art. Allowing the bidders to propose their own technical suggestions may give you information about different

options, thus helping you select the best methods to insure the technical quality of your project.

In addition, if you provide very specific technical requirements in the RFP, vendors' proposals may need only parrot the RFP, making it difficult for you to judge the depth of their technical comprehension and expertise without seeking additional information. If you do allow bidders to propose their own technical suggestions, you should require that bidders completely specify the approach and rationale for all major elements of their design. This information will help you compare approaches when reviewing the proposals and negotiate changes in proposed approaches with the vendor finally selected.

A disadvantage of allowing bidders to propose their own technical approaches (which may differ significantly) is that you may have a difficult time comparing value for cost.

Specifying in the RFP that the vendor will provide full technical reports lays the groundwork for later monitoring of the project. You may want to use technical "watchdogs" (experts) to review vendors' proposals or to review the vendor's technical work after the contract has been awarded. In the latter case, the expert functions as an outside consultant, hired by and working for you but paid for by the vendor. This strategy is particularly useful if advanced and/or complex statistical and technical procedures will be proposed by vendors. Technical assistance may be invaluable.

5. EXPECTED SERVICES AND PRODUCTS

Does the RFP address all expected services and products, and specify the quantity and quality of each that are expected? Does it specify which of them belong to whom?

As with other elements of the RFP, clarity and specificity are critical here. The cost of a project is often greatly affected by the number of tests and reports to be created, printed, delivered, and so forth. It is wise to address potential costs or savings of contract revisions. Sometimes vendors give "credit," which may be used in later phases of the same project. Obviously this is a poor arrangement if you will not be conducting business with this company in the future. In addition, the credit offered by a vendor may not adequately represent the true cost of the omitted work, thereby depriving the state of full value. In some situations you may be able to trade certain tasks or products for others as the project progresses and priorities change, or the vendor may reduce the final billing on a project.

The RFP should also specify exactly who will constitute the group(s) to be tested and how special populations (e.g., Spanish speakers, visually impaired) are to be addressed. For example, should special forms or special administration procedures be developed? RFPs should also specify which products will belong to the state and which to the vendor, so possible disputes can be avoided.

6. PERSONNEL LOADINGS

Does the RFP require bidders to justify personnel loadings by task and relevant qualifications? Does the state have veto power over proposed personnel and/or changes in critical personnel?

Sometimes it is important to have more than just the top people specified for a project. Changes in other key staff may have a profound effect, particularly if these people have some special expertise or knowledge of the project or related programs. You will want to use a "key personnel" clause to protect the state from both intentional ("bait and switch" tactics) and unintentional changes in personnel.

You may want to require bidders to provide a list of their proposed staff who are already committed to concurrent projects and bids in order to see the spread of key personnel should the vendor win other outstanding bids. You could also ask to be updated immediately prior to proposal review. It should be helpful to see the percent of key personnel's commitment to other projects on a monthly basis during the timeline of your project so that you can judge whether their availability will be adequate to your needs, especially at critical periods of your program. A task loading chart can help identify how serious the bidders are about various aspects of their proposal.

In fairness to vendors, each organization may be structured differently in regard to support resources, so the percent of time a

key person in one organization requires to meet contract specifications may differ significantly from the percent of time required by another. Level of experience may also impact the percent of time required to do a job. Since the appropriate commitment of key personnel can greatly influence the success of a project, checking the track record of the bidders may be one of the best ways to ensure adequate allocation of key resources.

7. BUDGET

Does the RFP request a budget at the task level? Does it specify a payment schedule or request that bidders propose one?

It is very important to provide a standard format for all bidders to use in presenting their budget proposals so that you can adequately compare their bids, particularly if little time is available for this comparison. If you must or want to select the lowest bidder, it is imperative that you be able to judge who is truly the lowest. A standard budget format should apply to both the summary page, which is very useful to reviewers, and to the details of the budget. If details are to be compared, it is important to have all bidders break the details down in the same way. If the project extends for more than one year, it is also helpful to have all bidders' budgets broken down by year at the same level of detail; the first year's budget should have the greatest detail.

A detailed budget is useful for a variety of reasons. It allows you to decide which services or products to omit if it is necessary to cut costs, or what credit to expect if part of a project is cancelled. Detail is particularly necessary if you allow variable costs in the proposals, but it can also be helpful with fixed costs. It allows you to check that all requested products and services appear in the bidder's budget and protects the state from problems that may result when something is inadvertently omitted from the RFP and/or the vendor's budget.

A caution: Excessive budgetary detail required by an RFP may drive away some potential bidders who do not find the effort worth their while. Be sure you can substantiate your need for all the figures you require in the proposals. The level of detail required ought to be in proportion to the size of the project, with greater detail for bigger projects.

8. QUALITY CONTROL & SCHEDULING

How can quality control and scheduling requirements be assured? Should there be penalties for failing to complete scheduled work on time?

The RFP can spell out deadlines for completing interim job tasks and can request interim or progress reports and drafts of final reports to be reviewed by the state office prior to the final version.

It also may be helpful to specify the turnaround time for state review of documents.

Requirements such as progress reports protect the state in the "worst case scenario" and may be relaxed in practice as is appropriate. Interim reports may be weekly, biweekly, monthly, or other. They usually summarize the work done and pending, critical decisions to be made by the state, and information to be provided by given dates. In addition to serving these important managerial functions, such reports can serve as documentation.

Many states include a paragraph in the RFP about penalties for failing to complete scheduled work on time or failure to meet certain other terms of the contract, usually in terms of a certain amount of money per day. A ceiling for penalties should be stated for the protection of both parties. This gives the state a time frame beyond which they can attempt to salvage a project by seeking alternate sources of services funded by the penalties from the contracted vendor. It gives vendors the ability to assess in advance the extent of financial risk involved. If no ceiling is given, some vendors may choose not to bid.

States sometimes require performance bonds to protect themselves when dealing with small bidders about whom they know little. A few states have required bid bonds. Vendors, especially small ones, say they tend not to go to this expense unless very motivated.

E. THE REVIEW PROCESS

This section discusses several aspects of the review process and the importance of pre-contract negotiations. It covers the following topics:

1. Criteria & Process
2. Weighting the Criteria
3. Reviewers
4. Oral Presentations
5. Pre-contract Negotiations

1. CRITERIA AND PROCESS

Are the selection criteria and review process clearly specified in the RFP?

Vendors need to know how their proposals will be judged. The more specific the criteria, the easier it is for the vendors to prepare their proposals and for the reviewers to pass judgment on them. Criteria are often stated in such general terms that it is difficult for a bidder to know just how the proposal will be judged.

States vary in the process they use to select a proposal. Many use a two-stage process in which they first rate the work plan and then look at the budget. A few do not even look at the budget if the work plan is not adequate. This approach avoids the problem of being tempted or forced to accept a "bargain" bid for a basically

inadequate proposal. However, other states argue that it is valuable to consider cost along with other criteria.

RFPs tend not to reveal the actual process by which reviewers come to a decision, such as whether they vote independently and then compute an average score for each proposal, use a consensus system, or use some other approach. Vendors, of course, would like as much information about the process as possible. However, in some states, testing directors cannot discuss areas controlled by general state bidding policies.

2. WEIGHTING THE CRITERIA

What weight should be accorded to the various components of the bid (e.g. technical merit, staff quality, corporate capability, and so forth)? How does the state balance technical quality and budget issues in making selection decisions?

Some states apportion a total of 100 points among the various criteria and then assign points to the proposals for each criterion. An advantage of this approach is that it indicates to the bidders the relative weight of the criteria. It is most helpful when the criteria themselves have been explicitly stated. For example, it is instructive to know that the state will give a certain number of points for proposals that provide a solution to a particular technical problem.

In states where the review process is strictly followed, it is important to state criteria carefully and to distinguish between "necessary" and merely "desired" attributes of a proposal. For example, a state once called for bidders to propose a creative approach to a task and then was forced to eliminate vendors who proposed good but not creative approaches because the proposals were technically "incomplete." Now this state indicates the necessary basics in the criteria section of its RFP, and after selecting a vendor, the state may request the use of desired creative techniques, for which it allows extra time and money. If the vendor is unable to do this, the state will subcontract part of the project to another company.

Vendors particularly appreciate candid information about the relative importance of technical merit and cost so that they can develop their proposals accordingly. If cost is an overriding factor, be straightforward about it. There is no sense in gathering technically grand proposals that cannot possibly be funded. When cost is particularly important, the vendors who know they cannot compete against the lowest cost operations will probably not submit bids, and they would prefer to make this decision than waste their efforts. If, on the other hand, technical considerations are very important (and cost is just "normally important" as opposed to "critical"), be candid about it. Avoid the false economy of implementing a bare bones plan that fails to provide needed quality. An efficient RFP process is to the state's advantage.

3. REVIEWERS

Who can/should serve as reviewers (internal and external), and what qualifications should they possess?

It is useful for the reviewers to represent a variety of perspectives. For example, a technical expert may be familiar with the vendors' level of technical expertise, and a school district representative has probably had to deal directly with the local consequences of good and bad work by various test scoring and reporting vendors. In some states the purchasing office has requirements regarding who can serve as a reviewer, what they may be paid, and whether/how you may be able to "train" them.

Unfortunately, it is possible that a reviewer may have a a grudge against a particular bidder and it may not be apparent until it is too late. Your best protection is to avoid persons who are overly opinionated or narrow-minded and to search for a balanced panel of reviewers. In addition, in order to validate or refute reviewers' input, you can obtain references for key personnel from the directors of previous projects on which the personnel have worked.

Reviewers are seldom trained despite the importance of their task, they sometimes serve without pay, and they are not usually held accountable for their decisions. To ameliorate the situation you may be able to discuss with potential reviewers what you hope to accomplish, who you expect will bid, what problems are to be solved,

and so forth, in order to "educate" the reviewers and to detect their biases ahead of time.

If possible, reviewers should have input into the development of the RFP. At the very least, they should be given a copy of the RFP, any previous proposals or planning documents, and any other information well in advance of reviewing the proposals.

Some states use an ongoing technical committee comprised of technical experts from school districts and universities around the state to review proposals and monitor the progress of programs over several years. This arrangement allows the committee members to feel ownership of the program and to remain involved; they are most likely to make responsible decisions in guiding the program. In order to avoid any hint of bias in the review process, people who join the committee sign an agreement not to be on retainer to vendors whose proposals the committee may review.

4. ORAL PRESENTATIONS

Will oral presentations be possible or required? Will they occur before or after a proposal is selected? Will there be other opportunities for clarification of bids?

Oral presentations can benefit both vendors and states. They can allow the vendors to more fully explain their proposals and explore issues or priorities with the state. Orals can, in turn, provide

the state with valuable details about the proposals and information about key personnel. If there is no clear winner among the proposals, oral presentations may help the reviewers select a contractor. It may also be useful to have the option to schedule an oral with the apparent winner of a contract before the final letting of the contract.

Since in many states the RFP and proposal become part of the contract, it is imperative to resolve any discrepancies between the two documents. The expense and effort involved if orals are mandatory may burden small, distant vendors, and some of them may choose not to bid, thus reducing the choices available to the state. This problem can be avoided by making the orals optional, allowing bidders the choice of oral or written clarification of their proposal. However, reviewers may strongly prefer to question bidders in person, particularly if the oral is to be with the final bidders or with an apparent winner.

It is important to recognize that a great deal of effort may be required ahead of time to prepare reviewers to be objective and to avoid being swayed by the slickness of some vendors' presentations.

5. PRE-CONTRACT NEGOTIATIONS

Are pre-contract negotiations referred to in the RFP as part of the contracting process? What factors should be negotiated?

Pre-contract negotiations and specifications of special conditions are invaluable tools for contracting agencies. After selecting a vendor to provide the required test services, and before or while the contract is drawn up, a critical period exists during which you should negotiate a number of important aspects of the project with the contractor. Although the RFP and proposal include explicit timelines, work plans, and so forth, reality may differ from good intentions. For example, the purchasing office may have taken longer than expected to accomplish its tasks, necessitating a compressed timeline if a given deadline is still to be met. You will want to clarify or negotiate the details of the actual timeline, plus the work plan, staff assignments, schedule of meetings, involvement of outside groups such as curriculum committees, number of objectives to be represented on the test, and so forth.

Some states use this pre-contract period for oral screening of the selected vendor before any award is made. This opportunity to examine needed changes provides a safety valve for both sides -- allowing vendors to change their minds or back out at the last minute, thereby averting larger problems down the line. Referring to this procedure in the RFP informs bidders of what to expect.

PART II

TOWARD BETTER TECHNICAL QUALITY IN LARGE SCALE ASSESSMENT PROGRAMS: ISSUES TO CONSIDER

A. Equating

B. Item Bias

C. Content Validity

A. ISSUES TO BE CONSIDERED IN THE EQUATING PORTIONS OF REQUESTS FOR PROPOSALS FOR LARGE SCALE ASSESSMENT PROGRAMS

Richard M. Jaeger
University of North Carolina at Greensboro

Introduction

Because of test security problems and the evolution of school curricula, large scale assessment programs require the creation of multiple forms of tests. For a variety of reasons--such as ensuring that each examinee has an equal opportunity to evidence his/her achievement, and assessing temporal trends in the average achievement in schools and school systems--it is essential that multiple forms of tests used in large scale assessments be placed on comparable score scales. The process used to effect scale comparability is termed "test equating."

Strictly speaking, tests that are to be equated must be psychometrically parallel. Frederic Lord has noted that two tests are equivalent only if it is a point of indifference to any examinee which test (s)he completes.

Test Equating Specifications

Since psychometric theory is replete with alternative methods for equating tests (cf. Angoff, 1984; Petersen, Kolen, and Hoover,in press) and none has been demonstrated to be universally superior, Requests for Proposals (RFPs) should specify a particular equating procedure only if the issuing state strongly prefers that procedure.

It is suggested that RFPs contain the following sections pertaining to test equating:

1. A detailed narrative on the purposes of test equating in the context of the statewide assessment program. Among several potential purposes—listed in order of increasing problems and difficulties—are:

a) Equating psychometrically parallel, multiple forms of a test;
b) Equating a slightly customized norm-referenced achievement test to a nationally normed standard form;
c) Equating a moderately customized norm-referenced achievement test to a nationally normed standard form;
d) Equating an extensively customized norm-referenced achievement test to a nationally normed standard form;

e) Equating a curriculum-tailored criterion-referenced test to a nationally standardized norm-referenced test; and

f) Placing multiple levels of a test intended for different grade levels or age levels of students on a continuous, longitudinally-interpretable scale;

2. Requirements that the proposal contain detailed discussion of the procedures to be used in equating tests or test forms to achieve each specified purpose. Among the procedures to be discussed should be:

a) The data-collection design to be used, including plans for sampling examinees and plans for administering tests or test forms to be equated to each sample of examinees;

b) The sizes and composition of samples of examinees to be drawn, including specification of the sampling frames to be used, the sampling units to be used, and backup sampling proposed to compensate for nonresponse; and

c) The analytic equating methods to be employed, including discussion of the use of anchor tests or items (if any), and the precise statistical procedures to be used in constructing a comparable score scale for all tests and forms to be equated;

3. Requirements that the proposal contain a detailed justification of the sampling and analysis methods proposed for each equating purpose, including reasons for selecting the proposed methods instead of viable alternatives.

4. Requirements that the proposal contain a detailed discussion of the methods to be used to evaluate the quality of the equatings that result from the data collected and the analytic procedures employed. In particular, the proposal should discuss methods to be used to estimate the degree of random equating error overall and at various points on the score scale. In situations where equating is to be applied to a sequence of tests over a period of years, methods to be used to estimate the resulting degree of scale drift should be described and justified.

5. Requirements for independent checking of the equating to verify its accuracy, appropriateness, and so forth.

Authors of RFPs should realize that the current state of measurement science does not support the use of test equating for many of the purposes listed under Point 1. In particular, it is widely known that test equating is not robust when applied to: (1) tests that differ substantially in content; (2) tests that differ substantially in difficulty or reliability; (3) tests that are targeted to groups of examinees that differ substantially in ability; and (4) tests that assess a multiplicity of constructs that are differentially sensitive to instruction. The greater the differences among tests in any of the

factors just listed, the weaker will be the generalization of equating results to populations that differ in composition from the equating sample.

Although previous research has shown that pre-equating of test items is generally not sufficient to ensure equivalent test forms in operational use, every attempt should be made to construct test forms that are as nearly parallel in content distribution and psychometric properties as is possible. Requirements for careful attention to parallelism should be specified in RFPs.

REFERENCES

Angoff, W.H. (1984). *Scales, Norms, and Equivalent Scores.* Princeton, NJ: Educational Testing Service.

Petersen, N.S., Kolen, M.J. & Hoover, H.D. (in press). Scaling, norming, and equating, in R.L. Linn (Ed.), *Educational Measurement.* Washington, DC: American Council of Education.

B. ISSUES TO BE CONSIDERED IN THE ITEM BIAS PORTIONS OF REQUESTS FOR PROPOSALS FOR LARGE SCALE ASSESSMENT PROGRAMS

H.D. Hoover
Iowa Testing Program
The University of Iowa

Introduction

The question of test bias as it applies to various social and cultural groups is a multifaceted one that differs somewhat for achievement tests, as opposed to aptitude or ability tests. In the United States, similarities in schools, language, and common culture transmitted through the mass media and population mobility make common nationwide or statewide achievement testing a meaningful endeavor. At the same time, it is also clear that there are significant curriculum differences among schools, that language differences exist across regions and cultural groups, and that the common culture is supplemented by many rich and unique cultural experiences. Some of the implications of this diversity for large scale statewide assessment programs are that:

(1) Tests should focus primarily on common experiences of all students;

(2) Special efforts must be made to avoid content unfamiliar to the experience of special groups and to balance familiarity of content for the various major cultures of the state or country;

(3) In interpreting test scores, emphasis should be placed on the individuality of pupils and the unique cultural circumstances that affect educational development; and

(4) Norms provided with the tests should adequately represent this cultural diversity.

Some methods that have been used by test publishers and other researchers to minimize cultural bias have included:

(1) Employing contributing test authors with diverse cultural backgrounds;

(2) Selecting materials that reflect the varied interests of pupils from a wide range of cultural backgrounds and experiences;

(3) Reviewing materials at all stages of preparation for unfairness of lack of relevance or unfairness for diverse groups;

(4) Conducting item tryouts in culturally diverse groups, analyzing results for potential item bias, and using this information in item selection and revision;

(5) Conducting research on relationships between cultural background and such factors as academic aptitude, achievement, social acceptance, persistence, and extracurricular participation; and

(6) Conducting research on educational and testing needs for different groups.

A distinction should be made between the potential bias that is a characteristic of the measuring instrument per se and bias resulting from the process of fallible human beings making decisions based, at least in part, on test evidence. Bias in test instruments may be more or less equated with lack of relevance. A test or test item which more nearly meets the individual needs of one pupil rather than another is less relevant for the latter and might be said to be biased against him or her. If differences in interest and motivation are considered to be biasing factors, all tests, or all experiences, may be said to have a certain amount of bias. A certain reading passage or language item might be more interesting and motivating for a girl than for a boy, for someone who is sports-minded, for someone from an urban environment as opposed to a rural environment, or for someone who is interested in science as opposed to someone who is interested in literature. A test that requires a pupil to do creative thinking is thus biased against a pupil who is not accustomed to thinking creatively. Examples of these subtly biased situations are all much easier to find than items that favor one ethnic group over another. Differences in motivation, interests, and values are extremely variable in all subcultures. This variability may explain in

part the low reliability exhibited by most statistical methods used to detect biased items (e.g., Hoover and Kolen, 1984).

Thus, in a sense, a given item or passage or even a whole test might be fairer for one pupil than another. If bias is defined in this way, it is difficult to conceive of a test that does not present some advantage for a given pupil or group of pupils. If all "bias" of this kind were to be removed, it would result in the elimination of all that is interesting, clever, novel, challengi.., and creative. Such a test would be bland, uninteresting, and irrelevant for everyone. (The attempt by textbook publishers to protect themselves from similar allegations of unfairness or "bias" is considered by many people to have been a major contributing factor in the "dumbing down" of textbooks).

Another situation sometimes cited as a potential source of bias is one that results from asking a given pupil a question based on something the pupil has never had an opportunity to learn. This could be a situation in which the knowledge is relevant, possibly even critical, but the school or society has not provided the opportunity to obtain it. One might reasonably contend that a test containing items of this type is not fair to the pupil, or even that the test is not valid in that it is not measuring what has been taught. However, if the purpose of testing is to improve instruction, it is exactly in this situation that a test has the potential for greatest usefulness because its use should lead to the provision of such experiences.

Minimizing the presence of bias in tests, of course, does not prevent the misuse of test results in decision making. The elimination of this type of bias requires wisdom and a deep-seated sense of personal responsibility on the part of all who use tests.

Responsibility for Fairness

The responsibility for insuring that tests used in large-scale assessment programs are as free from bias as is reasonably possible is one shared between the issuers of the RFP (hereafter referred to as the "state" for simplicity's sake) and the bidder or vendor. The degree of responsibility is related to the use and the ownership of the final test. As the ownership shifts to the state, so does the responsibility. Uses of the test not explicitly recommended by the publisher also shift responsibility to the state. The following examples should help clarify the nature of this shared responsibility:

1. An RFP calling for the use of a nationally standardized achievement battery (or shelf test) in the fall of the year where the testing program's primary focus is on the improvement of instruction, rather than accountability. In this case the responsibility would lie nearly totally with the vendor, since this use is one explicitly recommended for the test by the publisher. The vendor should also be able to furnish evidence pertaining to efforts used to ensure fairness, such

as the kinds of judgmental review and field testing of items described earlier. The state must judge from this evidence whether these efforts were satisfactory. In such situations little if any cost should be added to the vendor's bid.

2. An RFP calling for the use of a nationally standardized achievement battery to determine promotion from Grade 8 to Grade 9. Such a use is likely to impact various racial/ethnic groups differentially. In this case the responsibility should probably be shared roughly equally between the state and the vendor. Since such a use is not one normally recommended by the publisher, evidence in addition to that described in example 1 would be required. While the state might initially appear to have primary responsibility in this application, it should be kept in mind that the vendor is apt to benefit from the additional data obtained on the test battery pertaining to validity and "bias." The sharing of cost, along with the responsibility for fairness, would seem reasonable in this context.

3. An RFP requesting a criterion-referenced test that is to become the property of the state; a test tailored to its curriculum and intended for use in high stakes

decisions, similar to example 2. In this situation both the cost and the responsibility would lie predominantly with the state. The contracting of item review, sampling, and analysis procedures to a vendor would not abrogate this responsibility. The RFP should be quite explicit with respect to the methods to be used to insure equity in such an application. In fact, a separate RFP dealing only with item or test bias might be preferable.

These examples indicate that in some situations the responsibility for fairness lies nearly solely with the vendor and in others it lies more with the state. In those cases where the primary responsibility is the state's, it may still be reasonable for the state to contract this responsibility to the vendor. However, the procedures for the vendor to follow in these situations must be made explicit by the RFP.

If the test is used for high stakes decisions affecting individual students, the state takes more responsibility. Such use necessitates careful attention to item bias issues.

Item Bias Specifications in RFPs

There are two commonly used ways of screening potentially biased items from intact tests or from pools of test items. With judgmental methods, experts evaluate the fairness to various groups

of the item development process, of the presentation format, and of the content of potential items. With analytical methods, item data are obtained from relevant subgroups and indices sensitive to differential performance by these subgroups are computed. Judgmental methods are especially helpful in dealing with perceived fairness issues such as balance and unintended stereotyping. While they may also be of some help in minimizing differences in item performance among groups, most studies comparing judgmental and analytical methods have found the two to be essentially uncorrelated. For this reason, the use of test scores of individuals in high-stakes decision-making requires some attention to analytical methods.

Specific recommendations related to each of the two methods follow:

1) Judgmental. If a judgmental review of items is required, the RFP should document the process to be followed in item development to ensure fairness, provided the RFP requires "new" items. If shelf items are to be used, the RFP should ask for procedures used in item development. Any procedures required for content or linguistic review should also be made explicit in the RFP. If judges representative of specific racial/ethnic groups are expected to be part of this process it should be stated. However, it must be kept in mind that if the state expects shelf items from the vendor, highly restrictive specifications may preclude many potential bidders from responding.

2) Analytical. If empirical procedures requiring the use of "item bias indices" are included as a part of the RFP, a number of issues must be kept in mind. As is the case for equating procedures, a number of alternative item bias indices exist; none has been shown to be universally superior to the others. However, procedures utilizing only differences between groups in average percents correct have been shown in general to be inappropriate and should be avoided. Given adequate sample size, some of the procedures based on item response theory (IRT) appear promising. However, because of the unidimensionality assumption underlying IRT models, it is unresolved as to whether items identified by such methods as being "biased" might simply be indicating differences in dimensionality among groups (Linn and Harnisch, 1981).

In many states, a majority of the students of a given racial/ethnic group may be enrolled in a limited number of the school districts of that state. If curricula in these districts differ appreciably from those in the rest of the state, racial/ethnic differences in performance are nearly totally confounded with curriculum differences. It might be argued that tests, or items, measuring different things for different cultural groups are by definition biased. However, as was pointed out earlier, if the primary purpose of testing is to improve instruction, this is exactly the situation in which a test is most useful. In high-stakes testing applications (e.g., promotion or certification), it is especially important that data on possible curriculum and educational

background differences be obtained. Such information is sometimes referred to as "opportunity-to-learn" data.

Another major consideration in the use of item bias indices is their demonstrated low degree of reliability and subsequent low predictive validity. This low reliability becomes especially apparent when curriculum differences are controlled (Hoover and Kolen, 1984). In general, it is recommended that item bias indices not be the sole criteria for decisions regarding test item fairness, but that they be used in conjunction with other relevant information, including judgmental review of items.

If data on racial/ethnic differences in performance are to be obtained as a part of the item review process, the RFP should state: (1) which racial/ethnic or linguistic groups are to be sampled; (2) which sample design is to be used and what stratification variables will be furnished by the state; (3) whether opportunity-to-learn data is to be gathered (strongly recommended for high stakes decisions); and (4) which item bias methodology is preferred by the state.

The preceding discussion of issues and recommendations are primarily directed toward how to deal with item bias prior to the operational use of a test. It is recommended that states carefully analyze group differences in performance on intact tests and individual items after their first administration. This information should be used to improve tests for subsequent use. If it is expected

that these analyses are to be performed by the vendor, it should be stated in the RFP.

REFERENCES

Hoover, H. D. and Kolen, M. J. (1984). The reliability of six item bias indices. *Applied Psychological Measurement, 8*, 173-181.

Linn, R. L., and Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement, 18*, 109-118.

C. ISSUES TO BE CONSIDERED IN THE CONTENT VALIDITY PORTIONS OF REQUEST FOR PROPOSALS FOR LARGE SCALE ASSESSMENT PROGRAMS

Ron Hambleton
University of Massachusetts

According to the Standards for Educational and Psychological Testing (APA, 1985), content validity evidence requires reviewers to "assess the degree to which the sample of items in the test are representative of some defined domain of content." Expert judgment is the main mode of investigation of a test's content validity or related concepts, curricular validity, and instructional validity. The difference among the three approaches to validity determination is the particular domain of content to which test content is matched: In assessing content validity, test content is matched to the content specifications for the test; in assessing curricular validity, test content is matched to the domain of content defined by the curriculum of interest; and finally, in assessing instructional validity, test content is matched to what is actually taught in the curriculum in which the test will be used. Content validity evidence is important regardless of the test use. On the other hand, curricular validity evidence is most useful when choosing norm-referenced tests or conducting curriculum evaluation studies. Instructional validity evidence is especially useful with what are known as high-stakes tests (Popham, 1987).

In preparing content validity specifications for a Request for Proposal, the RFP writer generally has the choice of asking for a content validity plan, or providing specific details on the types of content validity evidence of interest:

a) Objective Representativeness. Are the objectives that are selected for inclusion in the test representative of all the objectives included in the domain of content?

b) Item Representativeness. Are the items that measure each objective in the test representative of the content that is defined by that objective?

c) Item-Objective Congruence. Does the item content seem appropriate to match the objectives of interest? Does successful performance on the test items require he same cognitive processes as those specified in the objectives the items were written to measure?

d) Technical Adequacy. Do the test items meet the standard item writing principles? Are the chosen item formats appropriate to permit valid assessment of the objectives of interest?

A few additional points concerning content validity studies are as follows:

a) Representativeness is assumed to mean assessing the more important or critical objectives, and reflecting the proportional size of the domains of content for objectives. In other words, for the representativeness criterion to be met, objectives that are more important or broader in scope than others need to be emphasized in test construction.

b) Judging "item or objective representativeness" may involve stratifying the domain of content prior to obtaining the reviewers' ratings.

Possible details to request from prospective contractors in an RFP include proposed selection and training of judges or reviewers, the number of judges, the intended review process and sample rating form, and reporting and use of content validity data.

Additional Research and Development Issues

1. Guidelines for knowing when content validity evidence is sufficient to support the intended use of the test scores would be helpful. (The particular test use and the feasibility of collecting the criterion data are important considerations.)

2. Guidelines for documenting (reporting) content validity evidence would be helpful.

3. More research is needed on the actual procedures for carrying out the four types of analyses described above (a-d) Content validity evidence is greatly valued, but the process of collecting the relevant data—unlike the standard-setting process, for example—appears to need further study.

REFERENCES

American Psychological Association (1985). Standards for educational and psychological testing. Washington, DC: APA.

Popham, W.J. (May 1987). The merits of measurement - driven instruction. Phi Delta Kappan. 679-682.